Averaging of Stochastic Differential Equations: Kurtz's Theorem revisited

Jessika Walter*, Christof Schütte†, Grigoris A. Pavliotis‡, Andrew Stuart‡

May 19, 2004

Abstract

This article is concerned with averaging techniques for stochastic differential equations with different temporal scales. We reformulate one of the theorems of Kurtz [4] in the sense that some abstract conditions of the theorem are replaced by others that allow for simpler validation for specific applications. Furthermore, the reformulated theorem is applied to two different problems: first, to a stochastic dynamical systems where a slow mode is alternately coupled to different fast modes but a stochastic switching process controls to which one; and second, to Langevin equations with different temporal scales as they appear in molecular dynamics and materials science applications.

1 Introduction

In complex system modelling, one often finds mathematical models that consist of differential equations with different temporal and spatial scales. As a consequence, mathematical techniques for the elimination of some of the smallest scales have achieved considerable attention in the last years; the derivation of reduced models by means of averaging or homogenization techniques, or keywords like multiscale modelling may serve as typical links to this discussion.

This article is concerned with averaging techniques for stochastic differential equations with two different temporal scales. The papers of Kurtz [4] and Papanicolaou [7] belong to the classical texts on this subject. For the more applied direction of the field, they still seem to be used as the main references to the theory. However, from the perspective of applications in

*Department of Mathematics and Computer Science, Free University Berlin, Germany.
Supported by the DFG proirity program "Analysis, Modeling and Simulation of Multiscale Problems" (SPP1095) and by the Marie Curie Fellowship Association.
†Department of Mathematics and Computer Science, Free University Berlin, Germany.
‡Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK.

complex systems modelling the theorems in these papers contain rather abstract conditions. In this article we will reformulate the main theorem of Kurtz from [4] in the sense that we will eliminate the (perhaps) most abstract conditions on ergodicity and on the range of the shifted generator of the process under consideration. This will be achieved by reference to some results of DAVIES [1] and LUMER & PHILLIPS [8].

Furthermore, we illustrate the possible usefulness of the reformulated theorem by considering two different problems that, both, originate from complex system modelling: In the first example, we consider a stochastic process where a slow variable at each instance is coupled to one of two fast variables but where a stochastic switching process controls the switches from one fast variable to the other. Systems like this occur in situations where some fast mode of a system rarely switches from one almost invariant set in its state space to another one such that the time scale of the switching is as slow as the slow modes of the system. This situation and its relations to molecular dynamics has been discussed in [10].
The second example is concerned with averaging of Langevin equations with different temporal scales as they are frequently used for modelling the dynamical behavior of large molecular systems [9, 11].

2 Kurtz Theorem Revisited

2.1 Background

We are given a time homogeneous Markov process $\{X_t\}_{t\in\mathbf{R}^+}$ on the state space $\mathbf{X}$ via its transition kernel $p : \mathbf{R}^+ \times \mathbf{X} \times \mathcal{B}(\mathbf{X}) \to [0,1]$ with

$$p(t,x,A) \;=\; \mathbf{P}[X_t \in A \;|X_{t=0} = x].$$

With the Markov process we may associate the family of *propagators* $P_t, t \geq 0$ acting on $L^1(\mathbf{X})$ according to the formula

$$\int_A P_t f(z)\mathrm{d}z \;=\; \int_{\mathbf{X}} f(x)p(t,x,A)\mathrm{d}x \qquad \text{all } A \in \mathcal{B}(\mathbf{X}). \tag{1}$$

By exploiting some properties of the transition kernel it is easily seen that P_t forms a *semigroup* wrt. time t, see e.g. [3]. The semigroup P_t describes the evolution of the distributions of the Markov process X_t. Namely, if we consider the process as beginning not at a given point $X_0 := X_{t=0} = x$ but rather a random point X_0 with distribution $\mu_0(\mathrm{d}x) = f(x)\mathrm{d}x : \mathbf{P}[X_0 \in A] = \mu_0(A)$, the distribution at time t will be $P_t f(x)\mathrm{d}x$.

The *infinitesimal generator* $\mathcal{L}$ of the semigroup is defined by the equality

$$\mathcal{L}f \;=\; \lim_{t\to 0+} \frac{P_t f - f}{t},$$

the domain $\mathcal{D}(\mathcal{L})$ of $\mathcal{L}$ being the set of $f \in L^1(\mathbf{X})$ for which the limit exists. It is evident that $\mathcal{L}$ is a linear operator from $\mathcal{D}(\mathcal{L})$ into $L^1 := L^1(\mathbf{X})$. It is not generally the case that $\mathcal{D}(\mathcal{L})$ equals L^1, but it is dense in the space

$$L = \{f \in L^1 : \lim_{t\to 0+} \|P_t f - f\| = 0\}.$$

We call P_t a *strongly continuous* semigroup if

$$\lim_{t\to 0+} P_t f = f \quad \text{for every } f \in L^1.$$

Thus, if $P_t : L^1(\mathbf{X}) \to L^1(\mathbf{X})$ is strongly continuous in L^1 then $\mathcal{D}(\mathcal{L})$ is dense in $L^1(\mathbf{X})$, i.e., $\overline{\mathcal{D}(\mathcal{L})} = L^1$.

For the systems we are interested in the infinitesimal generator of the semigroup P_t arises in connection with the *Fokker-Planck* equation

$$\partial_t f_t = \mathcal{L} f_t,$$

where the solution is

$$f_t = P_t f_0.$$

A probability density f_* is said to be *invariant* under the Markov process X_t if $P_t f_* = f_*$. In terms of the generator $\mathcal{L}$ we can express the invariance of a density $f_* \in \mathcal{D}(\mathcal{L})$ equivalently by $\mathcal{L} f_* = 0$. Thus, every density from the nullspace of $\mathcal{L}$, denoted by $\mathcal{N}(\mathcal{L})$, gives rise to an invariant density of the process X_t. A Borel set $E \subset \mathbf{X}$ is said to be *invariant* with respect to a positive operator P on $L^1(\mathbf{X})$ if for $f \in L^1(\mathbf{X})$ with $\mathrm{Supp}(f) \subset E$ we have $\mathrm{Supp}(Pf) \subset E$, where

$$\mathrm{Supp}(f) = \{x \in \mathbf{X} : f(x) \neq 0\}$$

and all statements about sets are taken modulo null sets. The semigroup P_t is said to be *irreducible* if the only sets which are invariant with respect to all P_t are $\emptyset$ and $\mathbf{X}$. As is shown in [1, Chapter 7], irreducibility of the semigroup P_t immediately implies $\dim\mathcal{N}(\mathcal{L}) \leq 1$. Furthermore, if P_t is assumed to be irreducible with $\dim\mathcal{N}(\mathcal{L}) = 1$ the unique invariant density f_* is strictly positive which means that

$$\int_E f_*(x)\mathrm{d}x > 0$$

for every Borel set E with positive Lebesgue-measure [1, Chapter 7].

2.2 Reformulation of Kurtz's Theorem

Throughout the section we fix the σ-finite measure space $(\mathbf{Z}, \mathrm{d}z)$. Suppose that P_t^ϵ is a strongly continuous contraction semigroup acting on the space

$L^1 := L^1(\mathbf{Z})$ and depending on a smallness parameter ϵ. Our basic assumption will be that its generator $\mathcal{L}^\epsilon$ can be decomposed into the sum of two generators:

$$\mathcal{L}^\epsilon = \frac{1}{\epsilon}\mathcal{L}_1 + \mathcal{L}_2. \tag{2}$$

We are interested in what happens to P_t^ϵ as $\epsilon \to 0$.

Suppose that the semigroup P_t^ϵ corresponds to a Markov process

$$(X^\epsilon(t), Y^\epsilon(t)) \in \mathbf{X} \times \mathbf{Y} = \mathbf{Z}$$

where X^ϵ denotes the slow mode and Y^ϵ the fast mode (the part $\mathcal{L}_1/\epsilon$ of the generator incorporates the forces acting on Y^ϵ and the time scale of the dynamical behavior of Y^ϵ is assumed to scale with ϵ). As an example for the origin of such processes we may consider dynamical systems of the following form:

$$\frac{\mathrm{d}}{\mathrm{d}t}X^\epsilon = f(X^\epsilon, Y^\epsilon, \xi) \tag{3}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}Y^\epsilon = \frac{1}{\epsilon}g(X^\epsilon, Y^\epsilon, \eta, \epsilon) \tag{4}$$

where ξ and η are time-dependent stochastic processes, and f and g are chosen such that the solution is a Markov process with generator of form (2) . If we freeze $Y^\epsilon(t) \equiv y$ on the RHS of (3) then the solution $X_y^\epsilon(t) := (X^\epsilon(t), y)$ of (3) is independent of ϵ and can be considered as a Markov process corresponding to the infinitesimal generator $\mathcal{L}_2^y := \mathcal{L}_2$. Here, the index indicates the coordinate that can be considered fixed, i.e., for y fixed $\mathcal{L}_2^y$ acts on $f(x,y)$ as a function of x alone. Thus, we have to distinguish between $\mathcal{D}(\mathcal{L}_2) \subset L^1(\mathbf{X} \times \mathbf{Y})$ and $\mathcal{D}(\mathcal{L}_2^y) \subset L^1(\mathbf{X} \times \{y\})$. We will simply identify $\mathbf{X} \times \{y\} = \mathbf{X}$ in the following. In the same way we relate the process $Y_x^\epsilon(t)$ (given as the solution of (4) for frozen $X^\epsilon \equiv x$) to the generator $(1/\epsilon)\mathcal{L}_1$ where $\mathcal{L}_1$ is denoted by $\mathcal{L}_1^x$ if we want to say that it acts on $f = f(x,y)$ as a function of y alone. Thereby we get a family of operators $\mathcal{L}_1^x$ acting on the y-direction for fixed x. Again, the domain of $\mathcal{L}_1^x$ is seen as a subset of $L^1(\mathbf{Y})$, whereas $\mathcal{L}_1$ is considered as operator acting on functions $f = f(x,y) \in L^1(\mathbf{X} \times \mathbf{Y})$.

A basic demand for the convergence of P_t^ϵ to a limiting semigroup P_t as $\epsilon \to 0$ is that the process Y_t^ϵ is ergodic in a sufficiently strong sense. This is related to some requirements for the generator family $\mathcal{L}_1^x$. More precisely, we have to demand that for every $x \in X$ there exists a strictly positive density $\mu_x \in L^1(\mathbf{Y})$ such that $\mathcal{L}_1^x \mu_x = 0$. For simplicity let us additionally assume that the corresponding propagator semigroup S_t^x is irreducible[1], thus

[1] A comment on the assumption of irreducibility is given in the next remark

$\dim \mathcal{N}(\mathcal{L}_1^x) = 1$ for every $x \in X$. Let us now define the projection operator Π on $L^1(\mathbf{X} \times \mathbf{Y})$ by

$$(\Pi f)(x,y) \;=\; \mu_x(y) \cdot \int_Y f(x,y)\mathrm{d}y. \tag{5}$$

Thus, Π maps every function $f \in L^1(\mathbf{X} \times \mathbf{Y})$ onto the space of functions which can be written in the form

$$f(x,y) \;=\; \hat{f}(x) \cdot \mu_x(y),$$

where $\hat{f}$ is an arbitrary function of $L^1(\mathbf{X})$. Again, by fixing x the operator Π can be considered as acting on $L^1(\mathbf{Y})$. If this is meant we will write Π_x instead of Π, thus $\Pi_x : L^1(\mathbf{Y}) \to \mathcal{N}(\mathcal{L}_1^x)$. Now we are ready to present our theorem. The range of an operator A is denoted by $\mathcal{R}(A)$.

Theorem 2.1 *Let $P_t^\epsilon, \mathcal{L}^\epsilon, \mathcal{L}_1$, and $\mathcal{L}_2$ be defined as above. Furthermore assume that $\mathcal{L}_1$ and $\mathcal{L}_2$ are generators of strongly continuous contraction semigroups S_t and U_t, respectively. In addition, suppose that $\mathcal{L}_1$ is the closure of $\mathcal{L}_1$ restricted to $\mathcal{D}(\mathcal{L}_2) \cap \mathcal{D}(\mathcal{L}_1)$. For every $x \in \mathbf{X}$ suppose that S_t^x is irreducible and that there exists a strictly positive density $\mu_x \in L^1(\mathbf{Y})$ such that $\mathcal{L}_1\mu_x = \mathcal{L}_1^x\mu_x = 0$. Let Π denote the projection operator according to (5) and let $D = \mathcal{R}(\Pi) \cap \mathcal{D}(\mathcal{L}_2)$, and define $\bar{\mathcal{L}}_\mu$ by*

$$\bar{\mathcal{L}}_\mu f \;=\; \Pi\mathcal{L}_2 f \qquad \textit{for all } f \in D.$$

Suppose that the closure of $\bar{\mathcal{L}}_\mu$ is the infinitesimal generator of a strongly continuous semigroup P_t defined on $\overline{D}$ with $\overline{D}$ denoting the closure of D. Then the following property holds:

$$\lim_{\epsilon \to 0} P_t^\epsilon f \;=\; P_t f \qquad \textit{for all } f \in \overline{D}.$$

Proof: The proof is based on a Theorem of Kurtz [4] and on results by Davies [1], stated in the appendix as Theorem A.2 and Theorem A.3. In order to apply Theorem A.2 we have to show the following conditions:

(i)
$$\lim_{\lambda \to 0} \lambda \int_0^\infty e^{-\lambda t} S_t f \mathrm{d}t \;=\; \Pi f,$$
for all $f \in L^1(\mathbf{X} \times \mathbf{Y})$ and Π defined by (5).

(ii)
$$D \subset \overline{\mathcal{R}(\lambda - \bar{\mathcal{L}}_\mu)} \qquad \text{for some } \lambda > 0.$$

(i) can be verified by using Theorem A.3 in the appendix which has to be applied to the semigroup S_t^x for every $x \in X$: Take $f = f(x,y) \in L^1(\mathbf{X} \times \mathbf{Y})$ and fix the variable x such that $f_x := f(x,\cdot) \in L^1(\mathbf{Y})$. Now we apply Theorem A.3 to S_t^x which is assumed to be irreducible with $S_t^x \mu_x = \mu_x$. S_t^x obviously satisfies (22). Thus, we immediately get

$$\lim_{\lambda \to 0} \lambda \int_0^\infty e^{-\lambda t} S_t f(x,\cdot) \mathrm{d}t \;=\; \lim_{\lambda \to 0} \lambda \int_0^\infty e^{-\lambda t} S_t^x f_x \mathrm{d}t \;=\; \Pi_x f_x, \tag{6}$$

where the limit is in the sense of strong convergence in $L^1(\mathbf{Y})$ for fixed x. Let us define F_λ by

$$F_\lambda(x,y) := \lambda \int_0^\infty e^{-\lambda t} S_t f(x,\cdot) \mathrm{d}t - \Pi f(x,y).$$

We now have to show that $F_\lambda(x,y)$ converges to 0 in $L^1(\mathbf{X} \times \mathbf{Y})$, i.e.,

$$\int_x \int_y |F_\lambda(x,y)| \mathrm{d}x \mathrm{d}y \;\to\; 0, \qquad \text{as } \lambda \to 0. \tag{7}$$

According to (6) we know that

$$\tilde{F}_\lambda(x) := \int_y |F_\lambda(x,y)| \mathrm{d}y \;\to\; 0$$

pointwise for every $x \in X$ as $\lambda \to 0$. Suppose that there exists an integrable function $\tilde{F} \in L^1(\mathbf{X})$ such that

$$\tilde{F}_\lambda(x) \;\leq\; |\tilde{F}(x)| \tag{8}$$

for every $x \in X$. Then we are able to apply Lebesgue's dominated convergence theorem to get the desired convergence (7). Thus, we only have to show (8):

$$\begin{aligned}
\int_y |F_\lambda(x,y)| \mathrm{d}y \;&\leq\; \|\lambda \int_0^\infty e^{-\lambda t} S_t^x f_x \mathrm{d}t\|_{L^1(\mathbf{Y})} \;+\; \|\Pi_x f_x\|_{L^1(\mathbf{Y})} \\
&\leq\; \lambda \int_0^\infty e^{-\lambda t} \|S_t^x f_x\|_{L^1(\mathbf{Y})} \mathrm{d}t \;+\; \|\Pi_x f_x\|_{L^1(\mathbf{Y})} \\
&\leq\; \lambda \int_0^\infty e^{-\lambda t} \|f(x,\cdot)\|_{L^1(\mathbf{Y})} \mathrm{d}t \;+\; \|\Pi_x f_x\|_{L^1(\mathbf{Y})} \\
&\leq\; \|f(x,\cdot)\|_{L^1(\mathbf{Y})} \;+\; \|\Pi_x f_x\|_{L^1(\mathbf{Y})},
\end{aligned}$$

which is integrable in $L^1(\mathbf{X})$ as we have chosen $f \in L^1(\mathbf{X} \times \mathbf{Y})$.

For (ii) we observe that for all $\lambda > 0$ we have $\mathcal{R}(\lambda - \bar{\mathcal{L}}_\mu) = \overline{D}$ since we assumed that $\bar{\mathcal{L}}_\mu$ is the infinitesimal generator of a strongly continuous semigroup on $\overline{D}$. This is due to the Theorem of Lumer-Phillips which can be found in the appendix. $\square$

Remark 2.2 *The reformulated theorem no longer contains the conditions on ergodicity and on the range of the shifted generator. The new conditions on irreducibility of the fast process and on the existence of a strictly positive invariant measure are more easily checked, for example for systems that originate from statistical mechanics, molecular dynamics, or materials science.*

It is possible to formulate the theorem even if we do not assume irreducibility of the process but only the existence of a strictly positive invariant measure.

Theorem 2.3 *Let $P_t^\epsilon, \mathcal{L}^\epsilon, \mathcal{L}_1$, and $\mathcal{L}_2$ be defined as above. Furthermore assume that $\mathcal{L}_1$ and $\mathcal{L}_2$ are generators of strongly continuous contraction semigroups S_t and U_t, respectively. In addition, suppose that $\mathcal{L}_1$ is the closure of $\mathcal{L}_1$ restricted to $\mathcal{D}(\mathcal{L}_2) \cap \mathcal{D}(\mathcal{L}_1)$. For every $x \in \mathbf{X}$ suppose that there exists a strictly positive density $\mu_x \in L^1(\mathbf{Y})$ such that $\mathcal{L}_1 \mu_x = \mathcal{L}_1^x \mu_x = 0$. Let Π denote the projection operator defined by*

$$\Pi f := \lim_{\lambda \to 0} \lambda \int_0^\infty e^{-\lambda t} S_t f \,\mathrm{d}t$$

and let $D = \mathcal{R}(\Pi) \cap \mathcal{D}(\mathcal{L}_2)$. Define $\bar{\mathcal{L}}_\mu$ by

$$\bar{\mathcal{L}}_\mu f = \Pi \mathcal{L}_2 f \qquad \text{for all } f \in D.$$

Suppose that the closure of $\bar{\mathcal{L}}_\mu$ is the infinitesimal generator of a strongly continuous semigroup P_t defined on $\overline{D}$ with $\overline{D}$ denoting the closure of D. Then the following property holds:

$$\lim_{\epsilon \to 0} P_t^\epsilon f = P_t f \qquad \text{for all } f \in \overline{D}.$$

Remark 2.4 *Due to [1] the expression for the projection Π in Theorem 2.3 is equivalently given by the* mean ergodic projection

$$\Pi f := \lim_{r \to \infty} \frac{1}{r} \int_0^r S_t f \,\mathrm{d}t \quad \text{for each } f \in L^1(\mathbf{X} \times \mathbf{Y}).$$

For example, these results enable us to apply the Theorem of Kurtz even to the deterministic Hamiltonian system with slow and fast parts (resulting from so-called strong constraining potentials) which obviously is not irreducible. This is investigated in another project and will be part of a forthcoming paper

3 Example 1: SDE with Switching Process

Let us consider the following SDE with Markovian switching of the form

$$\begin{aligned} \dot{x}^\epsilon(t) &= g(x^\epsilon(t), y^\epsilon(t)) + \sigma \dot{W}, & (9) \\ \epsilon \dot{y}^\epsilon(t) &= -\delta_{s(t)} y^\epsilon(t) + \sqrt{\epsilon}\sigma_{s(t)} \dot{W}_{s(t)}, & (10) \end{aligned}$$

with $x, y \in \mathbf{R}$, $\epsilon > 0$ and $W, W_{s(t)}$ denoting standard Brownian motions. Let $s(t)$ be a right-continuous Markov chain on a probability space taking values in a finite state space $\mathbf{S} = \{1, 2, ..., N\}$ and δ_i, σ_i take values in $\mathbf{R}^+$ for all $i \in \mathbf{S}$. The generator $\mathcal{S} = (s_{ij})_{N\times N}$ of the Switching chain $s(t)$ is given by

$$\mathbf{P}[s(t+\mathrm{d}t) = j | s(t) = i] = \begin{cases} s_{ij}\mathrm{d}t + o(\mathrm{d}t) & \text{if } i \neq j, \\ 1 + s_{ij}\mathrm{d}t + o(\mathrm{d}t) & \text{if } i = j, \end{cases}$$

where $\mathrm{d}t > 0$. Here $s_{ij} > 0$ is the transition rate from i to j if $i \neq j$ while

$$s_{ii} = -\sum_{i \neq j} s_{ij}. \tag{11}$$

For $\epsilon \ll 1$, this system consists of a fast variable, y, and a slow one, x. For fixed $i \in \mathbf{S}$ we will denote the process in the fast variable y by y_i^ϵ. Each process y_i^ϵ is an Ornstein-Uhlenbeck process and consequently ergodic. In the following we denote the stationary density of each y_i^ϵ by μ_i, where μ_i is given by

$$\mu_i(y) = (1/\sigma_i)\sqrt{\delta_i/\pi}\exp(-\delta_i y^2/\sigma_i^2), \tag{12}$$

which is the Gaussian density with mean zero and variance $\sigma_i^2/2\delta_i$, and thus independent of ϵ.

Remark 3.1 *It is emphasized that one should not confuse the evolution of a single trajectory $y^\epsilon(t)$ with the dynamics given by*

$$\tilde{y}^\epsilon(t) = \mathbf{1}_{\{s(t)=1\}} y_1^\epsilon(t) + \mathbf{1}_{\{s(t)=2\}} y_2^\epsilon(t) + ... + \mathbf{1}_{\{s(t)=N\}} y_N^\epsilon(t)$$

where y_i^ϵ is the solution process of

$$\epsilon \dot{y}_i^\epsilon(t) = -\delta_i y(t) + \sqrt{\epsilon}\sigma_i \dot{W}_i, \qquad i = 1, 2, ...N. \tag{13}$$

The difference should be obvious: The trajectories $y_i^\epsilon(t)$, $i = 1, ..., N$, evolve independently of each other and the dynamics of $\tilde{y}^\epsilon$ is governed by cutting out time segments of $y_i^\epsilon(t), i \in \mathbf{S}$, according to the switching process $s(t)$, respectively. However, the trajectory $y^\epsilon(t)$ due to (10) is given by changing the dynamics due to the switching chain such that we do not get points of discontinuity at the jumping times of $s(t)$.

Systems like (9) & (10) occur in situations where some fast mode of a system rarely switches from one almost invariant set in its state space to another one such that the time scale of the switching is as slow as the slow modes of the system. This situation and its relations to molecular dynamics has been discussed in [10].

3.1 Propagator semigroup and Fokker-Planck equation

Let $Z^\epsilon(t)$ denote the $\mathbf{R}^2 \times \mathbf{S}$-valued process $(x^\epsilon(t), y^\epsilon(t), s(t))$. Then $Z^\epsilon(t)$ is a time homogeneous Markov process. Due to (1) the family of propagators $P_t^\epsilon, t \geq 0$ acting on $L^1(\mathbf{R}^2 \times \mathbf{S})$ is given by the formula

$$(P_t^\epsilon f)(\mathrm{d}z, \{j\}) \;=\; \sum_{i\in\mathbf{S}} \int_{\mathbf{R}^2} f(\tilde{z}, i)\, p^\epsilon(t, \tilde{z}, i, \mathrm{d}z, \{j\})\mathrm{d}\tilde{z},$$

where $p^\epsilon(t, \tilde{z}, i, \mathrm{d}z, \{j\})$ denotes the transition function corresponding to the Markov solution process $Z^\epsilon(t) = (x^\epsilon(t), y^\epsilon(t), s(t))$ of (9)&(10) for given initial condition $(x^\epsilon(0), y^\epsilon(0)) = \tilde{z}$, $s(0) = i$.

Using the notation of Section 2.2 we note that the processes $X^\epsilon(t)$, $Y^\epsilon(t)$ are given by $X^\epsilon(t) = (x^\epsilon(t), s(t))$, $Y^\epsilon(t) = y^\epsilon(t)$, thus $\mathbf{X} = \mathbf{R} \times \mathbf{S}$, $\mathbf{Y} = \mathbf{R}$. The infinitesimal generator $\mathcal{L}^\epsilon$ of the semigroup P_t^ϵ is given for every $f \in \mathcal{D}(\mathcal{L}^\epsilon)$ by the operator $\mathcal{L}^\epsilon f : \mathbf{R}^2 \times \mathbf{S} \to \mathbf{R}$ being defined by

$$\mathcal{L}^\epsilon f \;=\; \frac{1}{\epsilon}\mathcal{L}_1 f \,+\, \mathcal{L}_2 f$$

with

$$\begin{aligned}
\mathcal{L}_1 f(x,y,i) &= \frac{\sigma_i^2}{2}\Delta_y f(x,y,i) \,+\, \nabla_y(\delta_i y\, f(x,y,i)) \\
\mathcal{L}_2 f(x,y,i) &= \frac{\sigma^2}{2}\Delta_x f(x,y,i) - \nabla_x(g(x,y)f(x,y,i)) + \sum_{j\in\mathbf{S}} s_{ji}(x) f(x,y,j).
\end{aligned}$$

As $\mathcal{L}_1$ acts as a differential operator on the fast variable y only, it can be considered in the space $L^1(\mathbf{Y})$ as well. If this is the case, i.e., if we consider $\mathcal{L}_1$ as acting on functions of y only, we will denote it $\mathcal{L}_1^i$. In accordance with Section 2.2, the notation $\mathcal{L}_1^i$ is used to indicate the coordinates that can be considered fixed for the respective operation. As $\mathcal{L}_1$ does not depend on the variable x, we will write $\mathcal{L}_1^i$ instead of $\mathcal{L}_1^{x,i}$. Analogously, for fixed y the generator $\mathcal{L}_2^y := \mathcal{L}_2$ is defined for functions depending on x and i. Thus, it should be clear that

$$\begin{aligned}
\mathcal{D}(\mathcal{L}_1), \mathcal{D}(\mathcal{L}_2) &\subset L^1(\mathbf{X} \times \mathbf{Y}) = L^1(\mathbf{R}^2 \times \mathbf{S}), \\
\mathcal{D}(\mathcal{L}_1^i) &\subset L^1(\mathbf{Y}) = L^1(\mathbf{R}), \\
\mathcal{D}(\mathcal{L}_2^y) &\subset L^1(\mathbf{X}) = L^1(\mathbf{R} \times \mathbf{S}).
\end{aligned}$$

In order to assure strong continuity of the semigroup generated by $\mathcal{L}_2$ the function $g = g(x, y)$ from (9) has to be chosen appropriately. For example, this is fulfilled for $g = 0$ or $g = x$. Another example is discussed at the end of Section 3.2. Thus, we make the following assumption on the drift coefficient $g = g(x, y)$.

Assumption 3.2 *Suppose for each $y \in \mathbf{Y}$ that $\mathcal{L}_2^y$ generates a strongly continuous contraction semigroup U_t^y on $L^1(\mathbf{X})$, i.e., $\overline{\mathcal{D}(\mathcal{L}_2^y)} = L^1(\mathbf{X})$. Moreover, assume that $\cap_{y \in \mathbf{R}} \mathcal{D}(\mathcal{L}_2^y)$ is dense in $L^1(\mathbf{X})$ and that the closure of $\mathcal{L}^\epsilon = \mathcal{L}_1/\epsilon + \mathcal{L}_2$ generates a strongly continuous contraction semigroup P_t^ϵ.*

The strong continuity of U_t^y for every y on $L^1(\mathbf{X})$ implies that U_t is a strongly continuous semigroup on $L^1(\mathbf{X} \times \mathbf{Y})$, where

$$(U_t f)(x, y, i) = (U_t^y f_y)(x, i), \qquad f_y := f(\cdot, y, \cdot).$$

Note, that the density of $\cap_y \mathcal{D}(\mathcal{L}_2^y)$ in $L^1(\mathbf{X})$ implies $\mathcal{L}_1/\epsilon + \mathcal{L}_2$ is densely defined for all $\epsilon > 0$ and that $\mathcal{L}_1$ is the closure of $\mathcal{L}_1$ restricted to $\mathcal{D}(\mathcal{L}_2) \cap \mathcal{D}(\mathcal{L}_1)$. Thus, due to the Theorem of Lumer-Phillips the assumption that $\mathcal{R}(\lambda - (\mathcal{L}_1/\epsilon + \mathcal{L}_2))$ is dense in $L^1(\mathbf{X} \times \mathbf{Y})$ for some $\lambda > 0$ would already imply the strong continuity of P_t^ϵ on $L^1(\mathbf{X} \times \mathbf{Y})$.

Let us now consider the generator $\mathcal{L}_1^i$ of the Markov process defined by (13) which is known as the Ornstein-Uhlenbeck process. The evolution of densities is governed by the Fokker-Planck equation $\partial_t f = \mathcal{L}_1^i f$, $(\epsilon = 1)$. Due to Davies [2, Chapter 4.3], this defines a strongly continuous contraction semigroup $S_t^i = \exp(t\mathcal{L}_1^i)$ on $L^1(\mathbf{R})$. The following is known about the semigroup S_t^i (see e.g. [5, Chapter 11.7]):

(i) S_t^i is irreducible;

(ii) the semigroup possesses a (unique) invariant density μ_i that is given by (12).

3.2 Reduced System

We want to average with respect to the fast variable y and obtain an averaged equation for the slow variable x alone. Thus, we would like to derive an equation for the distribution function in x:

$$\sum_{i \in \mathbf{S}} \int \rho^\epsilon(t, x, y, i) \mathrm{d}y$$

which should be valid in the limit $\epsilon \to 0$. To this end, we introduce the density $\hat{\rho}^\epsilon$ with

$$\hat{\rho}^\epsilon(t, x, i) = \int \rho^\epsilon(t, x, y, i) \mathrm{d}y.$$

It is expected that an approximate solution of the full dynamics would be obtained by multiplying each $\hat{\rho}^\epsilon(t,x,i)$ by the stationary distribution $\mu_i(y)$ of the SDE (13). We formalize this by defining the *projection operator* Π acting on functions $f = f(x,y,i) \in L^1(\mathbf{R}^2 \times \mathbf{S})$ by

$$(\Pi f)(x,y,i) \;=\; \mu_i(y) \cdot \int f(x,y,i)\mathrm{d}y. \tag{14}$$

Under the Assumption 3.2 the results of Section 2.2 are applicable in this situation. To this end, we calculate the 'reduced' operator $\bar{\mathcal{L}}_\mu$ defined by $\bar{\mathcal{L}}_\mu := \Pi\mathcal{L}_2$ on the pre-set domain $D = \mathcal{R}(\Pi) \cap \mathcal{D}(\mathcal{L}_2)$. Now we observe that D can be expressed by

$$D \;=\; \{\mu_i(y) \cdot \hat{f}(x,i) : \quad \hat{f} \in \cap_{y\in\mathbf{R}}\mathcal{D}(\mathcal{L}_2^y)\}.$$

Thus, for $f(x,y,i) = \mu_i(y) \cdot \hat{f}(x,i)$ simple calculations reveal

$$(\bar{\mathcal{L}}_\mu f)(x,y,i) \;=\; \mu_i(y) \cdot \bar{\mathcal{L}}\hat{f}(x,i).$$

The operator $\bar{\mathcal{L}}$ is defined for $\hat{f} \in \cap_y\mathcal{D}(\mathcal{L}_2^y)$ by

$$\bar{\mathcal{L}}\hat{f}(x,i) \;=\; \frac{\sigma^2}{2}\Delta_x\hat{f}(x,i) - \nabla_x(G_i(x)\cdot\hat{f}(x,i)) + \sum_{j\in\mathbf{S}} s_{ji}(x)\hat{f}(x,j),$$

where for $i \in \mathbf{S}$

$$G_i(x) \;=\; \int g(x,y)\,\mu_i(y)\mathrm{d}y.$$

Assumption 3.3 *The functions G_i imply that the closure of $\bar{\mathcal{L}}$ generates a strongly continuous semigroup $e^{t\bar{\mathcal{L}}}$ on $L^1(\mathbf{X}) = L^1(\mathbf{R}\times\mathbf{S})$.*

Remark 3.4 *Let us suppose that for every $i \in \mathbf{S}$ there is a $y_i \in \mathbf{Y}$ such that*

$$G_i(x) \;=\; g(x,y_i), \qquad \text{for all } x \in \mathbf{R}.$$

Then $\bar{\mathcal{L}}\hat{f}(x,i) = \mathcal{L}_2^{y_i}\hat{f}(x,i)$ such that the domain of $\bar{\mathcal{L}}$ is given by

$$\mathcal{D}(\bar{\mathcal{L}}) = \cap_{i\in\mathbf{S}}\mathcal{D}(\mathcal{L}_2^{y_i}).$$

This implies that the closure of $\bar{\mathcal{L}}$ generates a strongly continuous semigroup $e^{t\bar{\mathcal{L}}}$ on $L^1(\mathbf{X}) = L^1(\mathbf{R}\times\mathbf{S})$ where its domain contains $\cap_y\mathcal{D}(\mathcal{L}_2^y)$.

With Assumption 3.3 being valid, Theorem 2.1 is applicable where the limiting semigroup P_t is generated by the closure of $\bar{\mathcal{L}}_\mu$ on D and is strongly continuous on

$$\bar{D} \;=\; \{\mu_i(y)\cdot\hat{f}(x,i) : \qquad \hat{f} \in L^1(\mathbf{R}\times\mathbf{S})\}.$$

Conclusively, the evolution of densities $f = \mu_i(y) \cdot \hat{f}(x,i) \in \bar{D}$ is governed by the semigroup $e^{t\bar{\mathcal{L}}}$ on $L^1(\mathbf{R} \times \mathbf{S})$ according to

$$(P_t f)(x,y,i) \;=\; \mu_i(y) \cdot (e^{t\bar{\mathcal{L}}}\hat{f})(x,i),$$

and Kurtz's Theorem yields weak convergence, i.e., strong convergence of densities:

$$\lim_{\epsilon \to 0} P_t^\epsilon f \;=\; P_t f \qquad \text{in } \bar{D}.$$

$\bar{\mathcal{L}}$ is the infinitesimal generator associated with the SDE

$$\begin{aligned} \dot{x}^0 &= G_{s(t)}(x^0) + \sigma \dot{W}, \\ G_i(x) &= \int g(x,y)\mu_i(y)\mathrm{d}y, \end{aligned}$$

with $s(t)$ being the Markovian switching process with transition rates given by (11).

Example. We briefly discuss the case where the drift term g is given by

$$g(x,y) \;=\; -x \cdot y^2,$$

which is the case if we introduce the potential $V(x,y) = \frac{1}{2} \cdot x^2 \cdot y^2$, i.e., $-\nabla_x V(x,y) = g(x,y)$. For fixed $y \neq 0$, it is known that $\mathcal{L}_2^y$ generates a strongly continuous semigroup where the domain contains $C_0(\mathbf{R} \times \mathbf{S})$ with C_0 denoting continuous functions vanishing at infinity, see e.g. [2]. For $y = 0$, the domain of $\mathcal{L}_2^{y=0}$ is given by the Schwartz space. Thus, we observe that $\cap_y \mathcal{D}(\mathcal{L}_2^y)$ is dense in $L^1(\mathbf{X})$.

The averaged functions G_i are given by

$$G_i(x) \;=\; -x \cdot \int_{\mathbf{R}} y^2 \mu_i(y)\mathrm{d}y \;=\; -x \cdot \frac{\sigma_i^2}{2\delta_i} \;=\; g(x, y_i)$$

with $\sigma_i^2/2\delta_i$ denoting the variance of the density μ_i, respectively, according to (12), and $y_i = \sqrt{\sigma_i^2/2\delta_i}$. The limiting generator $\bar{\mathcal{L}}$ corresponds to the averaged system

$$\dot{x}^0 = G_{s(t)}(x^0) + \sigma \dot{W},$$

and the domain of the generator is $\cap_{i \in \mathbf{S}} \mathcal{D}(\mathcal{L}_2^{y_i})$. Thus, Theorem 2.1 is applicable since Assumption 3.2 and Assumption 3.3 are fulfilled.

4 Example 2: Langevin equation

As a second example we consider the following SDE

$$\dot{x}^\epsilon = p^\epsilon, \qquad \dot{p}^\epsilon \;=\; -\nabla_x V - p^\epsilon + \sigma \dot{W}_1 \tag{15}$$

$$\epsilon \dot{y}^\epsilon = \nu^\epsilon, \qquad \epsilon \dot{\nu}^\epsilon \;=\; -\nabla_y V - \nu^\epsilon + \sqrt{\epsilon}\sigma \dot{W}_2. \tag{16}$$

with $\epsilon > 0$, and $\nabla_x V, \nabla_y V$ denote the derivatives of the potential $V = V(x,y)$ wrt. x, y, respectively, and W_j, $j = 1,2$ standard Brownian motions. For $\epsilon \ll 1$, this system consists of fast variables (y,ν), and slow ones (x,p). By this example we illustrate the type of equations for which the velocity of the fast motion depends on the slow variables. It is well-known that the SDE (15)&(16) has an *invariant measure* $\mu(\mathrm{d}x, \mathrm{d}y, \mathrm{d}p, \mathrm{d}\nu) = \mu(x,y,p,\nu)\mathrm{d}x\mathrm{d}y\mathrm{d}p\mathrm{d}\nu$ with smooth density:

$$\mu(x,y,p,\nu) \;=\; \frac{1}{Z}\exp(-\beta V(x,y))\exp(-\beta\frac{1}{2}(p^2+\nu^2)), \quad \beta \;=\; \frac{2}{\sigma^2},$$

where Z denotes the normalization constant. The above SDE can also be written in second order form as

$$\begin{aligned} \ddot{x}^\epsilon \;&=\; -\nabla_x V \;-\; \dot{x}^\epsilon \;+\; \sigma \dot{W}_1 && (17)\\ \epsilon^2 \ddot{y}^\epsilon \;&=\; -\nabla_y V \;-\; \epsilon \dot{y}^\epsilon \;+\; \sqrt{\epsilon}\sigma \dot{W}_2, && (18)\end{aligned}$$

For abbreviation we will frequently use the notation:

$$\mathbf{x} \;=\; (x,p), \quad \mathbf{y} \;=\; (y,\nu).$$

Thus, $\mathbf{x} \in \mathbf{X} = \mathbf{R}^2$ denotes the slow variables and $\mathbf{y} \in \mathbf{Y} = \mathbf{R}^2$ the fast degrees of freedom.

The Fokker-Planck forward equation associated with the SDEs (15)&(16) reads

$$\begin{aligned} \partial_t \rho^\epsilon \;&=\; \mathcal{L}^\epsilon \rho^\epsilon && (19)\\ \mathcal{L}^\epsilon \;&=\; \frac{1}{\epsilon}\mathcal{L}_1 \;+\; \mathcal{L}_2 \;=\; \frac{1}{\epsilon}\mathcal{L}_1^x \;+\; \mathcal{L}_2^y \\ \mathcal{L}_1^x \;&=\; \frac{\sigma^2}{2}\Delta_\nu \;+\; \nu\cdot\nabla_\nu \;+\; \nabla_y V(x,\cdot)\cdot\nabla_\nu \;-\; \nu\cdot\nabla_y \;+\; 1 \\ \mathcal{L}_2^y \;&=\; \frac{\sigma^2}{2}\Delta_p \;+\; p\cdot\nabla_p \;+\; \nabla_x V(y,\cdot)\cdot\nabla_p \;-\; p\cdot\nabla_x \;+\; 1 \end{aligned}$$

The indices of the operators $\mathcal{L}_1^x$ and $\mathcal{L}_2^y$ indicate the coordinate that can be considered fixed for the respective operation, e.g., $\mathcal{L}_1^x$ can be considered as a differential operator acting on $\mathbf{y} = (y,\nu)$, but depending on x via the potential function $V(x,\cdot)$, where for y fixed $\mathcal{L}_2^y$ acts on a function $f \in L^1(\mathbf{X}\times\mathbf{Y})$ as a function of $\mathbf{x} = (x,p)$ alone.

The Fokker-Planck equation (19) describes the evolution of the probability density ρ^ϵ under the dynamics given by (15)&(16) and the semigroup of propagators P_t^ϵ is generated by the closure of the operator $\mathcal{L}^\epsilon$. For simplicity we will assume the strong continuity of the semigroup generated by $\mathcal{L}^\epsilon$ on $L^1(\mathbf{X}\times\mathbf{Y})$. Furthermore, we will make the following assumptions:

(A1) $\mathcal{L}_1$ and $\mathcal{L}_2$ are generators of strongly continuous semigroups S_t, U_t, respectively, where their domains are considered as dense subsets of $L^1(\mathbf{X}\times\mathbf{Y})$;

(A2) $\mathcal{L}_1$ is the closure of $\mathcal{L}_1$ restricted to $\mathcal{D}(\mathcal{L}_2)\cap\mathcal{D}(\mathcal{L}_1)$.

4.1 Reduced System

By fixing the slow variable $x \in \mathbf{R}$ we introduce the random process $\mathbf{y}_x(t) = (y_x(t), \dot{y}_x(t))$, which is defined by the stochastic differential equation

$$\ddot{y}_x \;=\; -\nabla_y V(x,\cdot) \,-\, \dot{y}_x \,+\, \sigma \dot{W}_2.$$

The solutions of this equation form a Markov process in $\mathbf{Y} = \mathbf{R}^2$, depending on $x \in \mathbf{R}$ as a parameter. The evolution of densities is governed by the strongly continuous semigroup S_t^x with infinitesimal operator $\mathcal{L}_1^x$. Let us impose the following conditions on the potential $V(x,\cdot)$ for fixed $x \in \mathbf{R}$:

Proposition 4.1 *The function $V(x,\cdot) \in C^\infty$ satisfies*

1. *$V(x,y) \geq 0$ for all $y \in \mathbf{R}$;*
2. *$V(x,\cdot)$ is a polynomial growing at infinity like $\|y\|^{2l}$, with l a positive integer.*

Then, the semigroup S_t^x is irreducible and possesses a (unique) invariant density μ_x given by

$$\begin{aligned} \mu_x(y,\nu) &= \frac{1}{Z(x)} \exp\left(-\beta\, V(x,y)\right) \frac{\sqrt{\beta}}{\sqrt{2\pi}} \exp(-\beta \frac{1}{2}\nu^2) \\ Z(x) &= \int \exp\left(-\beta\, V(x,y)\right) \mathrm{d}y. \end{aligned}$$

Proof: The statements can easily be verified by using Theorem 3.2 and Lemma 3.4 of Mattingly et al. [6] □

Let us now return to the full system and define the projection operator Π for functions $f = f(x,p,y,\nu) = f(\mathbf{x},\mathbf{y}) \in L^1(\mathbf{X} \times \mathbf{Y})$ according to

$$\Pi f(\mathbf{x},\mathbf{y}) \;=\; \mu_x(\mathbf{y}) \cdot \int_{\mathbf{Y}} f(\mathbf{x},\mathbf{y}) \mathrm{d}\mathbf{y}.$$

For every $x \in \mathbf{R}$ the density μ_x is normalized to 1, i.e., $\|\mu_x\|_{L^1(\mathbf{Y})} = 1$. Thus, it is obvious that the range of Π is given by

$$\mathcal{R}(\Pi) \;=\; \{\mu_x(\mathbf{y}) \cdot \hat{f}(\mathbf{x}) : \qquad \hat{f} \in L^1(\mathbf{X})\}.$$

It follows, that every $f \in D = \mathcal{D}(\mathcal{L}_2) \cap \mathcal{R}(\Pi)$ can be written as $f = \mu_x(\mathbf{y}) \cdot \hat{f}(\mathbf{x})$, where we subsequently assume that $\hat{f} \in \cap_y \mathcal{D}(\mathcal{L}_2^y) \subset L^1(\mathbf{X})$. Since $\mathcal{L}_2$ contains derivatives wrt. $\mathbf{x} = (x,p)$ only, we immediately see that for $f = \mu_x(\mathbf{y}) \cdot \hat{f}(\mathbf{x})$

$$\bar{\mathcal{L}}_\mu f \;:=\; (\Pi \mathcal{L}_2 f)(\mathbf{x},\mathbf{y}) \;=\; \mu_x(\mathbf{y}) \cdot \bar{\mathcal{L}} \hat{f}(\mathbf{x}),$$

where $\bar{\mathcal{L}}$ is defined according to

$$\bar{\mathcal{L}} = \frac{\sigma^2}{2}\Delta_p + p\cdot\nabla_p + \underbrace{\int_{\mathbf{Y}} \nabla_x V(x,y)\cdot\mu_x(\mathbf{y})\,\mathrm{d}\mathbf{y}}_{=I.} \cdot\nabla_p - p\cdot\nabla_x + 1.$$

One easily computes that $I.$ can again be expressed as the gradient of the *averaged potential*, i.e.,

$$\int_{\mathbf{Y}} \nabla_x V(x,y)\cdot\mu_x(\mathbf{y})\,\mathrm{d}\mathbf{y} = \nabla_x\bar{V}(x)$$

with

$$\bar{V}(x) \;=\; -\frac{1}{\beta}\log Z(x),$$

where $Z(x)$ has been introduced before as the normalization constant of the reduced invariant measure on the phase space fiber of the fast variable. Therefore,

$$\bar{\mathcal{L}} \;=\; \frac{\sigma^2}{2}\Delta_p + p\cdot\nabla_p + \nabla_x\bar{V}\cdot\nabla_p - p\cdot\nabla_x + 1, \tag{20}$$

which is the infinitesimal generator of the semigroup $e^{t\bar{\mathcal{L}}}$ corresponding to the Langevin-equation

$$\ddot{x}^0 \;=\; -\nabla_x\bar{V}(x^0) - \dot{x}^0 + \sigma\dot{W}_1.$$

The invariant measure $\bar{\mu}$ of the reduced dynamics in the slow degrees of freedom is given by

$$\bar{\mu}(\mathbf{x}) \;=\; \int_{\mathbf{Y}} \mu(\mathbf{x},\mathbf{y})\,\mathrm{d}\mathbf{y},$$

equivalently expressed by

$$\bar{\mu}(x,p) = \frac{1}{\bar{Z}}Z(x)\exp(-\beta\frac{1}{2}p^2)$$

with $\bar{Z}$ denoting the normalization constant.

Theorem 2.1 is applicable if the closure of $\bar{\mathcal{L}}_\mu$ generates a strongly continuous semigroup P_t on $\bar{D}$, which is fulfilled whenever the closure of $\bar{\mathcal{L}}$ generates a strongly continuous semigroup on $L^1(\mathbf{X})$ denoted by $e^{t\bar{\mathcal{L}}}$. Note that every $f\in\bar{D}$ is written as $f=\mu_x(\mathbf{y})\cdot\hat{f}(\mathbf{x})$ with $\hat{f}\in L^1(\mathbf{X})$ such that

$$(P_tf)(\mathbf{x},\mathbf{y}) \;=\; \mu_x(\mathbf{y})\cdot(e^{t\bar{\mathcal{L}}}\hat{f})(\mathbf{x}).$$

Theorem 2.1 implies that for every $f\in\bar{D}$

$$\lim_{\epsilon\to 0} P_t^\epsilon f \;=\; P_t f,$$

where the limit is considered in the sense of strong convergence in L^1.

References

[1] E. B. Davies. *One–Parameter Semigroups.* Academic Press, London, 1980.

[2] E. B. Davies. *Heat Kernels and Spectral Theory.* Cambridge Uni. Press, Cambridge, 1989.

[3] W. Huisinga. *Metastability of Markovian systems: A transfer operator based approach in application to molecular dynamics.* PhD thesis, Free University Berlin, 2001.

[4] T. G. Kurtz. A limit theorem for perturbed operator semigroups with applications to random evolutions. *J. Funct. Anal.*, 12:55–67, 1973.

[5] A. Lasota and M. C. Mackey. *Chaos, Fractals and Noise*, volume 97 of *Applied Mathematical Sciences.* Springer, New York, 2nd edition, 1994.

[6] J. Mattingly, A. M. Stuart, and D. J. Higham. Ergodicity for SDEs and approximations: Locally Lipschitz vector fields and degenerated noise. *To appear Stoch. Proc. Appl.*, 2002.

[7] G. Papanicolaou. Some probabilistic problems and methods in singular perturbation. *Rocky Mountain Math. J.*, 6, 1976.

[8] A. Pazy. *Semigroups of Linear Operators and Applications to Partial Differential Equations.* Springer-Verlag, New York, Berlin, Heidelberg, Tokyo, 1983.

[9] G. Ramachandran and T. Schlick. Solvent effects on supercoiled DNA explored by Langevin dynamics simulation. *Phys. Rev. E*, 51:6188–6203, 1995.

[10] C. Schütte, J. Walter, C. Hartmann, and W. Huisinga. An averaging principle for fast degrees of freedom exhibiting long-term correlations. *submitted to SIAM Multiscale Modeling and Simulation, presently available via www.math.fu-berlin.de/∼biocomp*, 2003.

[11] G. Zhang and T. Schlick. The langevin/implicit-euler/normal-mode scheme (lin) for molecular dynamics at large timesteps. *J. Chem. Phys.*, 101:4995–5012, 1994.

A Appendix

Theorem A.1 *Let S_t be a strongly continuous semigroup on L with infinitesimal operator $\mathcal{L}_1$. Suppose*

$$\lim_{\lambda \to 0} \lambda \int_0^\infty e^{-\lambda t} S_t f \mathrm{d}t = \Pi f \tag{21}$$

exists for every $f \in L$. Then

1. *Π is a bounded linear projection, i.e., $\Pi^2 = \Pi$;*
2. *$S_t \Pi = \Pi S_t = \Pi$ all $t > 0$;*
3. *$\mathcal{R}(\Pi) = \mathcal{N}(\mathcal{L}_1)$(the null space of $\mathcal{L}_1$);*
4. *$\mathcal{R}(\mathcal{L}_1)$ is dense in $\mathcal{N}(\Pi)$;*
5. *$\mathcal{L}_1 \Pi f = 0$ all $f \in L$, $\Pi \mathcal{L}_1 f = 0$ all $f \in \mathcal{D}(\mathcal{L}_1)$.*

Let U_t and S_t be strongly continuous semigroups of linear contractions on a Banach space L with infinitesimal operators $\mathcal{L}_2$ and $\mathcal{L}_1$, respectively. Suppose that for each sufficiently small ϵ, the closure of $(1/\epsilon)\mathcal{L}_1 + \mathcal{L}_2$ is the infinitesimal operator of a strongly continuous semigroup P_t^ϵ on L. In addition, assume that $\mathcal{L}_1$ is the closure of $\mathcal{L}_1$ restricted to $\mathcal{D}(\mathcal{L}_2) \cap \mathcal{D}(\mathcal{L}_1)$. We are interested in what happens to P_t^ϵ as ϵ goes to zero.

Theorem A.2 (Kurtz) *Let U_t, S_t and P_t^ϵ be defined as above. Suppose S_t satisfies the conditions of Theorem A.1. Let*

$$D = \{f \in \mathcal{R}(\Pi) : f \in \mathcal{D}(\mathcal{L}_2)\},$$

and define $\bar{\mathcal{L}} f = \Pi \mathcal{L}_2 f$ for $f \in D$. Suppose $\overline{\mathcal{R}(\lambda - \bar{\mathcal{L}})} \supset D$ for some $\lambda > 0$. Then the closure of $\bar{\mathcal{L}}$ restricted so that $\bar{\mathcal{L}} f \in \bar{D}$ is the infinitesimal operator of a strongly continuous contraction semigroup P_t defined on $\bar{D}$ and $\lim_{\epsilon \to 0} P_t^\epsilon f = P_t f$ for all $f \in \bar{D}$.

Theorem A.3 *Let S_t be a strongly continuous semigroup of positive contractions on $L = L^1(Y, \mathrm{d}y)$ where $(Y, \mathrm{d}y)$ is a measure space. Suppose there exists a strictly positive $f_0 \in \mathcal{L}_1$ such that $\mathcal{L}_1 f_0 = 0$ with $\mathcal{L}_1$ being the generator of S_t. Then*

$$\lim_{\lambda \to 0} \lambda \int_0^\infty e^{-\lambda t} S_t f \mathrm{d}t = \Pi f$$

exists for all $f \in L$. If in addition S_t is irreducible and satisfies

$$\int_Y (S_t f)(y) \mathrm{d}y = \int_{\mathbf{Y}} f(y) \mathrm{d}y \tag{22}$$

for all $f \in L$ and $t \geq 0$, then

$$\Pi f = f_0 \int_{\mathbf{Y}} f(y) \mathrm{d}y$$

for all $f \in L$.

The next theorem gives a characterization of the infinitesimal generator of a strongly continuous semigroup ([8, Chapter 1]). We need some preliminaries.

Let L be a Banach space and let L^* be its dual. The value of $g \in L^*$ at $f \in L$ is denoted by $\langle g, f \rangle$ or $\langle f, g \rangle$. For every $f \in L$ the duality set $F(f) \subset L^*$ is defined by

$$F(f) = \{g \in L^* : \quad \langle g, f \rangle = \|f\|^2 = \|g\|^2\}.$$

From the Hahn-Banach theorem it follows that $F(f) \neq \emptyset$ for every $f \in L$.

Definition A.4 *A linear operator A is dissipative if for every $f \in \mathcal{D}(A)$ there is a $g \in L^*$ such that* $\mathrm{Re}\langle Af, g\rangle \le 0$.

Theorem A.5 (Lumer-Phillips) *Let A be a linear operator with dense domain $\mathcal{D}(A)$ in the Banach space L.*

(i) If A is dissipative and there is a $\lambda > 0$ such that the range, $\mathcal{R}(\lambda - A)$, of $\lambda - A$ is L, then A is the infinitesimal generator of a strongly continuous semigroup of contractions on L.

(ii) If A is the infinitesimal generator of a strongly continuous semigroup of contractions on L then $\mathcal{R}(\lambda - A) = L$ for all $\lambda > 0$ and A is dissipative. Moreover, for every $f \in \mathcal{D}(A)$ and every $g \in F(f)$, $\mathrm{Re}\langle Af, g\rangle \le 0$.